

University of Groningen

## Robust radio interferometric calibration using the t-distribution

Kazemi, S.; Yatawatta, S.

*Published in:*  
Monthly Notices of the Royal Astronomical Society

*DOI:*  
[10.1093/mnras/stt1347](https://doi.org/10.1093/mnras/stt1347)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2013

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Kazemi, S., & Yatawatta, S. (2013). Robust radio interferometric calibration using the t-distribution. *Monthly Notices of the Royal Astronomical Society*, 435(1), 597-605. <https://doi.org/10.1093/mnras/stt1347>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Robust radio interferometric calibration using the $t$ -distribution

S. Kazemi<sup>1</sup>★ and S. Yatawatta<sup>2</sup>

<sup>1</sup>*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*

<sup>2</sup>*ASTRON, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands*

Accepted 2013 July 16. Received 2013 July 12; in original form 2013 June 12

## ABSTRACT

A major stage of radio interferometric data processing is calibration or the estimation of systematic errors in the data and the correction for such errors. A stochastic error (noise) model is assumed, and in most cases, this underlying model is assumed to be Gaussian. However, outliers in the data due to interference or due to errors in the sky model would have adverse effects on processing based on a Gaussian noise model. Most of the shortcomings of calibration such as the loss in flux or coherence, and the appearance of spurious sources, could be attributed to the deviations of the underlying noise model. In this paper, we propose to improve the robustness of calibration by using a noise model based on Student's  $t$ -distribution. Student's  $t$ -noise is a special case of Gaussian noise when the variance is unknown. Unlike Gaussian-noise-model-based calibration, traditional least-squares minimization would not directly extend to a case when we have a Student's  $t$ -noise model. Therefore, we use a variant of the expectation–maximization algorithm, called the expectation–conditional maximization either algorithm, when we have a Student's  $t$ -noise model and use the Levenberg–Marquardt algorithm in the maximization step. We give simulation results to show the robustness of the proposed calibration method as opposed to traditional Gaussian-noise-model-based calibration, especially in preserving the flux of weaker sources that are not included in the calibration model.

**Key words:** instrumentation: interferometers – methods: numerical – methods: statistical – techniques: interferometric.

## 1 INTRODUCTION

Radio interferometry gives an enhanced view of the sky, with increased sensitivity and higher resolution. There is a trend towards using phased arrays as the building blocks of radio telescopes (LOFAR<sup>1</sup> and SKA<sup>2</sup>) as opposed to the traditional dish-based interferometers. In order to reach the true potential of such telescopes, calibration is essential. Calibration refers to the estimation of systematic errors introduced by the instrument (such as the beam shape and receiver gain) and also by the propagation path (such as the ionosphere), and correction for such errors, before any imaging is done. Conventionally, calibration is done by observing a known celestial object (called the external calibrator), in addition to the part of the sky being observed. This is improved by self-calibration (Cornwell & Wilkinson 1981), which is essentially using the observed sky itself for the calibration. Therefore, self-calibration entails consideration of both the sky as well as the instrument as unknowns. By iteratively refining the sky and the instrument model, the quality of

the calibration is improved by orders of magnitude in comparison to using an external calibrator.

From a signal processing perspective, calibration is essentially the maximum likelihood (ML) estimation of the instrument and sky parameters. An in-depth overview of existing calibration techniques from an estimation perspective can be found in, e.g., Boonstra & van der Veen (2003), van der Veen, Leshem & Boonstra (2004), van der Tol, Jeffs & van der Veen (2007) and Kazemi et al. (2011). All such calibration techniques are based on a Gaussian noise model, and the ML estimate is obtained by minimizing the least-squares cost function using a non-linear optimization technique such as the Levenberg–Marquardt (LM; Levenberg 1944; Marquardt 1963) algorithm. Despite the obvious advantages of self-calibration, there are also some limitations. For instance, Cornwell & Fomalont (1999) give a detailed overview of the practical problems in self-calibration, in particular due to errors in the initial sky model. It is well known that the sources not included in the sky model have lower flux (or loss of coherence), and Martí-Vidal et al. (2010) is a recent study on this topic. Moreover, under certain situations, fake or spurious sources could appear due to calibration as studied by Martí-Vidal & Marcaide (2008).

In this paper, we propose to improve the robustness of calibration by assuming a Student's  $t$ -noise model (Gosset 1908) instead of a

★E-mail: kazemi@astro.rug.nl

<sup>1</sup> The Low Frequency Array.

<sup>2</sup> The Square Kilometre Array.

Gaussian noise model. One of the earliest attempts in deviating from a Gaussian-noise-model-based calibration can be found in Schwab (1982), where instead of minimizing a least-squares cost function, an  $l_1$  norm minimization was considered. Minimizing the  $l_1$  norm is equivalent to having a noise model which has a Laplacian distribution (Aravkin, Friedlander & van Leeuwen 2012). The motivation for Schwab (1982) to deviate from the Gaussian noise model was the ever present outliers in the radio interferometric data.

In a typical radio interferometric observation, there is a multitude of causes for outliers in the data.

(i) Radio frequency interference caused by man-made radio signals is a persistent cause of outliers in the data. However, data affected by such interference are removed before any calibration is performed by flagging (e.g. Offringa et al. 2010). But there might be faint interference still present in the data, even after flagging.

(ii) The initial sky model used in self-calibration is almost always different from the true sky that is observed. Such model errors also create outliers in the data. This is especially significant when we observe a part of the sky that has sources with complicated, extended structure. Moreover, during calibration, only the brightest sources are normally included in the sky model and the weaker sources act together to create outliers.

(iii) During day-time observations, the Sun could act as a source of interference, especially during high solar activity. In addition the Galactic plane also affects the signals on short baselines.

(iv) An interferometer made of phased arrays will have sidelobes that change with time and frequency. It is possible that a strong celestial source, faraway from the part of the sky being observed, will pass through such a sidelobe. This will also create outliers in the data.

To summarize, model errors of the sky as well as the instrument will create outliers in the data and in some situations calibration based on a Gaussian noise model will fail to perform satisfactorily. In this paper, we consider the specific problem of the effect of unmodelled sources in the sky during calibration. We consider ‘robustness’ to be the preservation of the fluxes of the unmodelled sources. Therefore, our prime focus is to minimize the loss of flux or coherence of unmodelled sources, and our previous work (Yatawatta, Kazemi & Zaroubi 2012) has measured robustness in terms of the quality of calibration.

Robust data modelling using Student’s  $t$ -distribution has been applied in many diverse areas of research, and Lange, Little & Tylor (1989), Bartkowiak (2007) and Aravkin et al. (2012) are few such examples. However, the traditional least-squares minimization is not directly applicable when we have a non-Gaussian noise model, and we apply the expectation–maximization (EM; Dempster, Laird & Rubin 1977) algorithm to convert calibration into an iteratively reweighted least-squares minimization problem, as proposed by Lange et al. (1989). In fact, we use an extension of the EM algorithm called the expectation–conditional maximization either (ECME) algorithm (Liu & Rubin 1995) to convert calibration to a tractable minimization problem. However, we emphasize that we do not force a non-Gaussian noise model on to the data. In case there are no outliers and the noise is actually Gaussian, our algorithms would work as traditional calibration does.

The rest of the paper is organized as follows. In Section 2, we give an overview of radio interferometric calibration. We consider the effect of unmodelled sources in the sky during calibration in Section 3. Next, in Section 4, we discuss the application of Student’s  $t$ -noise model in calibration. We also present the weighted LM routine used in calibration. In Section 5, we present simulation results to show

the superiority of the proposed calibration approach in minimizing the loss in coherence and present conclusions in Section 6.

*Notation.* Lowercase bold letters refer to column vectors (e.g.  $\mathbf{y}$ ). Uppercase bold letters refer to matrices (e.g.  $\mathbf{C}$ ). Unless otherwise stated, all parameters are complex numbers. The matrix inverse, transpose, Hermitian transpose and conjugation are referred to as  $(\cdot)^{-1}$ ,  $(\cdot)^T$ ,  $(\cdot)^H$  and  $(\cdot)^*$ , respectively. The matrix Kronecker product is given by  $\otimes$ . The statistical expectation operator is given as  $E\{\cdot\}$ . The vectorized representation of a matrix is given by  $\text{vec}(\cdot)$ . The  $i$ th diagonal element of matrix  $\mathbf{A}$  is given by  $\mathbf{A}_{ii}$ . The  $i$ th element of a vector  $\mathbf{y}$  is given by  $y_i$ . The identity matrix is given by  $\mathbf{I}$ . Estimated parameters are denoted by a hat,  $\hat{(\cdot)}$ . All logarithms are to the base  $e$ , unless stated otherwise. The  $l_2$  norm is given by  $\|\cdot\|$  and the infinity norm is given by  $\|\cdot\|_\infty$ . A random variable  $X$  that has a distribution  $\mathcal{P}$  is denoted by  $X \sim \mathcal{P}$ .

## 2 DATA MODEL

We give a brief overview of radio interferometry in this section. For more information about radio interferometry, the reader is referred to Thompson, Moran & Swenson (2001) and to Hamaker, Bregman & Sault (1996) for the data model in particular. We consider the radio frequency sky to be composed of discrete sources, faraway from the Earth such that the approaching radiation from each one of them appears to be plane waves. We decompose the contribution from the  $i$ th source into two orthogonal polarizations  $\mathbf{u}_i = [u_{xi} \ u_{yi}]^T$ . The interferometric array consists of  $R$  receiving elements with dual polarized feeds. At the  $p$ th station, this plane wave causes an induced voltage, which is dependent on the beam attenuation as well as the radio frequency receiver chain attenuation. The induced voltages at the  $x$  and  $y$  polarization feeds  $\tilde{\mathbf{v}}_{pi} = [v_{xpi} \ v_{ypi}]^T$  due to a source  $i$  are given as

$$\tilde{\mathbf{v}}_{pi} = \mathbf{J}_{pi} \mathbf{u}_i. \quad (1)$$

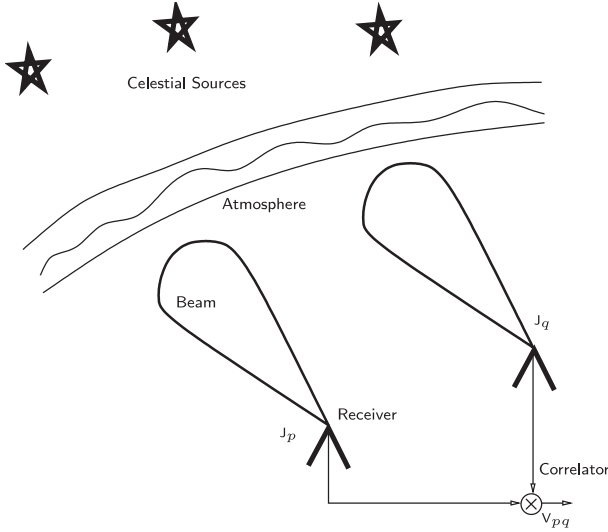
The  $2 \times 2$  Jones matrix  $\mathbf{J}_{pi}$  in equation (1) represents the effects of the propagation medium, the beam shape and the receiver. If there are  $K$  known sources (that are in the sky model) and  $K'$  unknown sources, the total signal will be a superposition of  $K + K'$  such signals as in equation (1).

Consider the correlation of signals at the  $p$ th receiver and the  $q$ th receiver, as shown in Fig. 1, with proper signal delay. After correlation, the correlated signal of the  $p$ th station and the  $q$ th station (named as the *visibilities*),  $\mathbf{V}_{pq} = E\{\mathbf{v}_p \mathbf{v}_q^H\}$ , is given by

$$\mathbf{V}_{pq} = \sum_{i=1}^K \mathbf{J}_{pi} \mathbf{C}_{pqi} \mathbf{J}_{qi}^H + \sum_{i'=1}^{K'} \mathbf{J}_{pi'} \mathbf{C}_{pqi'} \mathbf{J}_{qi'}^H + \mathbf{N}_{pq}. \quad (2)$$

In equation (2),  $\mathbf{J}_{pi}$  and  $\mathbf{J}_{qi}$  are the Jones matrices describing errors along the direction of source  $i$ , at stations  $p$  and  $q$ , respectively. The  $2 \times 2$  noise matrix is given as  $\mathbf{N}_{pq}$ . The contribution from the  $i$ th source on the baseline  $pq$  is given by the  $2 \times 2$  matrix  $\mathbf{C}_{pqi}$ . The noise matrix  $\mathbf{N}_{pq}$  is assumed to have elements with zero mean, complex Gaussian entries with equal variance in real and imaginary parts. Moreover, in equation (2), we have split the contribution from the sky into two parts:  $K$  sources that are known to us and  $K'$  sources that are unknown. Generally, the bright sources are always known but there are infinitely many faint sources that are too weak to be detected and too numerous to be included in the sky model. Therefore, almost always  $K'$  is much larger than  $K$ .

During calibration, we only estimate the Jones matrices  $\mathbf{J}_{pi}$  for  $p \in [1, R]$  and  $i \in [1, K]$ ; in other words, we estimate the errors along



**Figure 1.** A basic radio interferometer that correlates the signals received from faraway celestial sources. The signals are corrupted by Earth's atmosphere as well as by the receiver beam pattern, and these corruptions are represented by  $\mathbf{J}_p$  and  $\mathbf{J}_q$ .

the known bright sources. Due to our ignorance of the  $K'$  unknown sources, the effective noise during calibration becomes

$$\mathbf{N}'_{pq} = \sum_{i'=1}^{K'} \mathbf{J}_{pi'} \mathbf{C}_{pqi'} \mathbf{J}_{qi'}^H + \mathbf{N}_{pq} \quad (3)$$

and our assumption regarding the noise being complex circular Gaussian breaks down, depending on the properties of the signals of the weak sources. The prime motivation of this paper is to address this problem of the possible non-Gaussianity of the noise due to an error in the sky model. A similar situation could arise even for calibration along one direction (or direction independent calibration), when  $K = 1$ , if there is an error in the source model, for instance in the shape of the source.

The vectorized form of equation (2),  $\mathbf{v}_{pq} = \text{vec}(\mathbf{V}_{pq})$ , can be written as

$$\mathbf{v}_{pq} = \sum_{i=1}^K \mathbf{J}_{qi}^* \otimes \mathbf{J}_{pi} \text{vec}(\mathbf{C}_{pqi}) + \sum_{i'=1}^{K'} \mathbf{J}_{qi'}^* \otimes \mathbf{J}_{pi'} \text{vec}(\mathbf{C}_{pqi'}) + \mathbf{n}_{pq}, \quad (4)$$

where  $\mathbf{n}_{pq} = \text{vec}(\mathbf{N}_{pq})$ . Depending on the time and frequency interval within which calibration solutions are obtained, we can stack up all cross-correlations within that interval as

$$\mathbf{d} = [\text{real}(\mathbf{v}_{12}^T) \text{ imag}(\mathbf{v}_{12}^T) \text{ real}(\mathbf{v}_{13}^T) \dots \text{imag}(\mathbf{v}_{(R-1)R}^T)]^T, \quad (5)$$

where  $\mathbf{d}$  is a vector of size  $N \times 1$  of real data points. Thereafter, we can rewrite the data model as

$$\mathbf{d} = \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}) + \sum_{i'=1}^{K'} \mathbf{s}_{i'} + \mathbf{n}, \quad (6)$$

where  $\boldsymbol{\theta}$  is the real parameter vector (size  $M \times 1$ ) that is estimated by calibration. The contribution of the  $i$ th known source on all data points is given by  $\mathbf{s}_i(\boldsymbol{\theta})$  (size  $N \times 1$ ) and the unknown contribution from the  $i'$ th unknown source is given by  $\mathbf{s}_{i'}$  (size  $N \times 1$ ). The noise vector based on a Gaussian noise model is given by  $\mathbf{n}$  (size  $N \times 1$ ). The parameters  $\boldsymbol{\theta}$  are the elements of  $\mathbf{J}_{pi}$ s, with real and imaginary parts considered separately.

The ML estimate of  $\boldsymbol{\theta}$  under a zero mean, white Gaussian noise is obtained by minimizing the least-squares cost

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\| \mathbf{d} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}) \right\|^2 \quad (7)$$

as done in current calibration approaches (Boonstra & van der Veen 2003; van der Veen et al. 2004; van der Tol et al. 2007; Kazemi et al. 2011). However, due to the unmodelled sources, the effective noise is actually

$$\mathbf{n}' = \sum_{i'=1}^{K'} \mathbf{s}_{i'} + \mathbf{n} \quad (8)$$

even when  $\mathbf{n}$  is assumed to be Gaussian. Therefore, traditional calibration based on a least-squares cost minimization would not perform optimally. In order to improve this, we have to consider the statistical properties of the effective noise  $\mathbf{n}'$ , and we shall do that in Section 3.

### 3 EFFECT OF UNMODELLED SOURCES IN CALIBRATION

In this section, we study the effect of unmodelled sources on  $\mathbf{N}'_{pq}$  in equation (3) when  $\mathbf{N}_{pq}$  has elements with zero mean, complex circular white Gaussian statistics. We only select one element from the  $2 \times 2$  matrix (say at 1st row and column) for simplicity. Let us denote the baseline coordinates as  $u, v, w$  in wavelengths (we omit the  $pq$  subscript for simplicity). We can rewrite equation (3) for just one element as

$$z_{pq} = \sum_{i'=1}^{K'} g_{pqi'} I_{pqi'} \exp(-j2\pi(u l_{i'} + v m_{i'} + w(n_{i'} - 1))) + n_{pq}. \quad (9)$$

In equation (9),  $g_{pqi'}$  correspond to the corruptions along the direction  $i'$  (contributions from  $\mathbf{J}_{pi'}$  and  $\mathbf{J}_{qi'}$ ). The intensity of the  $i'$ th source seen on the baseline  $pq$  is given by  $I_{pqi'}$ . The direction cosines of the  $i'$ th source are given by  $l_{i'}$ ,  $m_{i'}$  and  $n_{i'}$ . The Gaussian noise is given by  $n_{pq} \sim \mathcal{CN}(0, \rho^2)$ . We assume that  $g_{pqi'}$ ,  $I_{pqi'}$ ,  $l_{i'}$ ,  $m_{i'}$ ,  $n_{i'}$  and  $n_{pq}$  are statistically independent from each other. Moreover, the sources are assumed to be uniformly distributed in a field of view defined by  $-\bar{l} \leq l_{i'} \leq \bar{l}$  and  $-\bar{m} \leq m_{i'} \leq \bar{m}$ , and that  $(n_{i'} - 1) \approx 0$ . The sources outside this area in the sky have almost no contribution to the signal due to the fact that the values of  $|g_{pqi'}|$  are very small, mainly due to attenuation by the beam shape.

With the above assumptions, we see that

$$E\{z_{pq}\} = \sum_{i'=1}^{K'} E\{g_{pqi'}\} E\{I_{pqi'}\} \text{sinc}(2\pi u \bar{l}) \text{sinc}(2\pi v \bar{m}) + E\{n_{pq}\} \quad (10)$$

which is almost zero if  $|u| > \frac{1}{2\bar{l}}$  and  $|v| > \frac{1}{2\bar{m}}$ . Therefore, we make the following assumptions applicable to long baselines:

- (i) the mean of the effective noise is almost equal to the mean of noise,  $E\{z_{pq}\} \rightarrow E\{n_{pq}\} = 0$ ,
- (ii) the variance of effective noise is greater than the variance of noise,  $E\{|z_{pq}|^2\} > E\{|n_{pq}|^2\}$ .

Let us briefly consider the implications of equation (10) above. First, the field of view is  $2\bar{l} \times 2\bar{m}$  in the sky. Now, consider the longest baseline length or the maximum value of  $\sqrt{u^2 + v^2}$  to be  $\bar{u}$ . Therefore, the image resolution will be about  $1/\bar{u}$ , and consider the

field of view to be of width  $2\bar{l} \approx P \times 1/\bar{u}$ . In other words, the field of view is  $P$  image pixels when the pixel width is  $1/\bar{u}$ . Now, in order for  $E\{z_{pq}\} \approx 0$  in equation (10), we need  $|u| > \frac{1}{2\bar{l}}$  or  $|u| > \bar{u}/P$  (and a similar expression can be derived for  $|v|$ ). This means that for baselines that are at least  $1/P$  times the maximum baseline length, we can assume  $E\{z_{pq}\} \approx 0$ .

To illustrate the above discussion, we give a numerical example considering the LOFAR highband array at 150 MHz. The point spread function at this frequency is about 6 arcsec and the field of view is about  $10^\circ$  in diameter. Therefore,  $P \approx 10 \times 3600/6 = 6000$ . The longest baselines is about 80 km and for all baselines that are greater than  $80/6 = 13$  m, the assumptions made above more or less hold.

To summarize the discussion in this section, we claim that

$$E \left\{ \sum_{i'=1}^{K'} s_{i'} \right\} \rightarrow \mathbf{0} \quad (11)$$

in equation (8) and, therefore,  $E\{\mathbf{n}'\} \rightarrow E\{\mathbf{n}\}$ . However, the covariance of  $\mathbf{n}'$  is different than the covariance of  $\mathbf{n}$ , and in general, the effective noise is not necessarily Gaussian anymore.

### 3.1 SAGE algorithm with unmodelled sources

In our previous work (Kazemi et al. 2011), we have presented the Space Alternating Generalized EM (SAGE; Fessler & Hero 1994) algorithm as an efficient and accurate method to solve equation (7), when the noise model is Gaussian. However, when there are unmodelled sources, as we have seen in this section, the noise model is not necessarily Gaussian.

The SAGE expectation step is finding the conditional mean of the  $k$ th signal,

$$\mathbf{x}^k = \mathbf{s}_k(\boldsymbol{\theta}) + \mathbf{n}' = \mathbf{s}_k(\boldsymbol{\theta}) + \sum_{i'=1}^{K'} s_{i'} + \mathbf{n}, \quad (12)$$

where  $\mathbf{x}^k$  is the hidden data. Using this, we can rewrite the observed data  $\mathbf{d}$  as

$$\mathbf{d} = \mathbf{x}^k + \sum_{i=1, i \neq k}^K s_i(\boldsymbol{\theta}). \quad (13)$$

The conditional mean of  $\mathbf{x}^k$  given  $\mathbf{d}$ , is given as  $\hat{\mathbf{x}}^k$ , where

$$\hat{\mathbf{x}}^k = \mathbf{s}_k(\boldsymbol{\theta}) + \left( \mathbf{d} - \sum_{i=1, i \neq k}^K s_i(\boldsymbol{\theta}) - \sum_{i'=1}^{K'} s_{i'} \right), \quad (14)$$

where we still assume a Gaussian noise model  $\mathbf{n}$ . Under the assumption that  $\sum_{i'=1}^{K'} s_{i'} \rightarrow \mathbf{0}$ , the conditional mean simplifies to

$$\hat{\mathbf{x}}^k \approx \mathbf{s}_k(\boldsymbol{\theta}) + \left( \mathbf{d} - \sum_{i=1, i \neq k}^K s_i(\boldsymbol{\theta}) \right). \quad (15)$$

The SAGE maximization step maximizes the likelihood of the conditional mean  $\hat{\mathbf{x}}^k$  under the noise  $\mathbf{n}'$ . However, we cannot use a least-squares cost function as  $\mathbf{n}'$  is not necessarily Gaussian anymore, because of the unmodelled sources. In Section 4, we explore an alternative noise model based on Student's  $t$ -distribution (Gosset 1908) for the maximization of the likelihood.

## 4 ROBUST CALIBRATION

First, we briefly describe the univariate Student's  $t$ -distribution (Lange et al. 1989; Bartkowiak 2007). Let  $X$  be a random variable with a normal distribution  $\mathcal{N}(\varepsilon, \sigma^2/\gamma)$ , where  $\gamma$  is also a random variable. Then the conditional distribution of  $X$  is

$$p(x|\varepsilon, \sigma^2, \gamma) = \frac{1}{(\sigma/\sqrt{\gamma})\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{x - \varepsilon}{\sigma/\sqrt{\gamma}} \right)^2 \right). \quad (16)$$

We consider  $\gamma$  to have a gamma distribution,  $\gamma \sim \text{gamma}(\frac{\nu}{2}, \frac{\nu}{2})$ , where  $\nu$  is positive (also called the number of degrees of freedom). The density function of  $\gamma$  can be given as

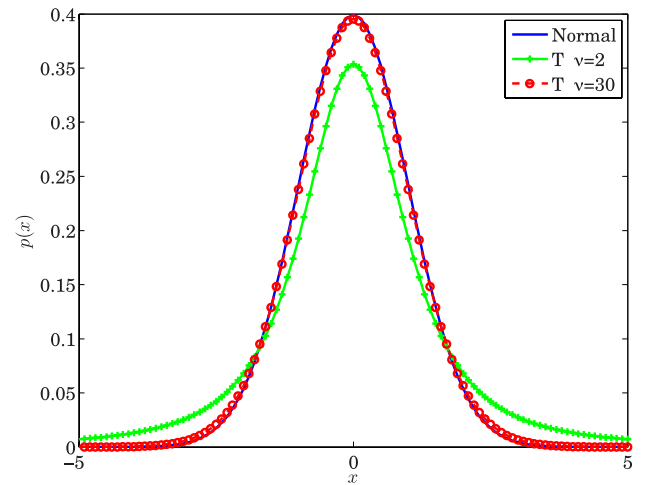
$$p(\gamma|\nu) = \frac{1}{\Gamma(\frac{\nu}{2})} \left( \frac{\nu}{2} \right)^{\frac{\nu}{2}} \gamma^{\frac{\nu}{2}-1} \exp \left( -\frac{\nu\gamma}{2} \right). \quad (17)$$

Then, the marginal distribution of  $X$  is

$$p(x; \varepsilon, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{(\pi\nu)^{1/2} \Gamma(\frac{\nu}{2}) \sigma} \left( 1 + \frac{1}{\nu} \left( \frac{x - \varepsilon}{\sigma} \right)^2 \right)^{-\frac{1}{2}(\nu+1)}, \quad (18)$$

and this is the probability density function which defines the Student's  $t$ -distribution. In Fig. 2, we have shown the probability density functions for Gaussian distribution and Student's  $t$ -distribution, both with zero mean and unit variance. We see that for low values of the number of degrees of freedom  $\nu$ , Student's  $t$ -distribution has a higher tail. The asymptotic limit of Student's  $t$ -distribution is Gaussian as  $\nu \rightarrow \infty$ , and for  $\nu > 30$ , the two distributions are indistinguishable, within the resolution of the data points used in Fig. 2.

Reverting back to equation (8), we see that the increase in the noise variance due to the unmodelled sources can be considered as the effect of  $\gamma$  in equation (16). Therefore, we consider the noise vector  $\mathbf{n}'$  to have independent, identically distributed entries, with the distribution given by equation (18) with  $\varepsilon = 0$  and  $\sigma = \rho = 1$ . In the SAGE iterations outlined in Section 3.1, at the  $k$ th iteration (15), we have  $\hat{\mathbf{x}}^k$  as the data vector and  $\mathbf{s}_k(\boldsymbol{\theta})$  as the model that is used to estimate the parameters  $\boldsymbol{\theta}$  (or a subset of the parameters). Therefore, the estimation problem is to find the ML estimate of  $\boldsymbol{\theta}$  (size  $M \times 1$ ), given the data  $\mathbf{y} = \hat{\mathbf{x}}^k$  (size  $N \times 1$ ) and the model



**Figure 2.** Probability density functions for standard normal distribution and Student's  $t$ -distribution, with  $\nu = 2$  and  $30$ . At  $\nu = 30$ , the Student's distribution is indistinguishable from the normal distribution.



$\mathbf{f}(\boldsymbol{\theta}) = \mathbf{s}_k(\boldsymbol{\theta})$  (size  $N \times 1$ ) with noise  $\mathbf{n}'$ . Hence, we can rewrite our data model as

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{n}', \quad (19)$$

where the unknowns are  $\boldsymbol{\theta}$  and  $\nu$ , the number of degrees of freedom of noise  $\mathbf{n}'$ . Then, the  $i$ th element of the vector  $\mathbf{y}$  (denoted by  $y_i$ ) in equation (6) will have a similar distribution as equation (18) with  $\sigma = 1$  and  $\mu_i = \mathbf{f}_i(\boldsymbol{\theta})$ , where  $\mathbf{f}_i(\boldsymbol{\theta})$  is the  $i$ th element of the vector function  $\mathbf{f}(\boldsymbol{\theta})$ . The likelihood function becomes

$$l(\boldsymbol{\theta}, \nu | \mathbf{y}) = \prod_{i=1}^N \frac{\Gamma(\frac{\nu+1}{2})}{(\pi\nu)^{1/2} \Gamma(\frac{\nu}{2})} \left( 1 + \frac{(y_i - \mathbf{f}_i(\boldsymbol{\theta}))^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)} \quad (20)$$

and the log-likelihood function is

$$L(\boldsymbol{\theta}, \nu | \mathbf{y}) = N \log \Gamma\left(\frac{\nu+1}{2}\right) - N \log \Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2} \log(\pi\nu) - \frac{(\nu+1)}{2} \sum_{i=1}^N \log\left(1 + \frac{(y_i - \mathbf{f}_i(\boldsymbol{\theta}))^2}{\nu}\right). \quad (21)$$

Note that unlike for the Gaussian case, minimizing a least-squares cost function (or maximizing the likelihood) will not give us the ML estimate. In addition, we have an extra parameter,  $\nu$ , which is the number of degrees of freedom. Hence, we use the ECME algorithm (Liu & Rubin 1995; Li, Wang & Chai 2006) to solve this problem. The ECME algorithm is an extension of the EM algorithm for  $t$ -distribution presented by Lange et al. (1989).

The auxiliary variables are the weights  $w_i$  ( $N$  values) and a scalar  $\lambda$ . All these are initialized to 1 at the beginning. The expectation step in the ECME algorithm involves the conditional estimation of hidden variables  $\gamma_i$  (or the weights  $w_i$ ) as

$$w_i \leftarrow E\{\gamma_i | y_i, \boldsymbol{\theta}, \nu\} = \lambda \frac{\nu+1}{\nu + (y_i - \mathbf{f}_i(\boldsymbol{\theta}))^2} \quad (22)$$

and the update of the scalar  $\lambda$

$$\lambda \leftarrow \frac{1}{N} \sum_{i=1}^N w_i. \quad (23)$$

The maximization step involves finding the value for  $\nu$  that is a solution for

$$\Psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{\nu+1}{2}\right) - \Psi(\nu/2) + \log(\nu/2) + \frac{1}{N} \sum_{i=1}^N (\log(w_i) - w_i) + 1 = 0, \quad (24)$$

where  $\Psi(x) = \frac{d}{dx} \log(\Gamma(x))$  is the digamma function. Since we know that beyond  $\nu > 30$  we almost get a Gaussian distribution, and therefore, the search space for finding a solution for equation (24) is kept within  $2 \leq \nu \leq 30$  and the initial value for  $\nu$  is chosen to be 2.

Once  $w_i$  is known,  $y_i$  has a normal distribution with variance determined by  $w_i$ . Therefore, in the maximization step of the EM algorithm, we minimize the weighted least-squares cost function:

$$l(\boldsymbol{\theta} | \nu) = \sum_{i=1}^N w_i (y_i - \mathbf{f}_i(\boldsymbol{\theta}))^2. \quad (25)$$

With this formulation at hand, we present the LM algorithm for robust calibration in Algorithm 1, similar to the presentations in Lourakis (2004) and Madsen, Nielsen & Tingleff (2004). The additional information needed in Algorithm 1 is the Jacobian of  $\mathbf{f}(\boldsymbol{\theta})$ ,

i.e.  $\mathbf{J}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ , that can be calculated in closed form using equations (2) and (4). The diagonal matrix with the weights  $\sqrt{w_i}$  as its diagonal entries is given by  $\mathbf{W}$ .

**Algorithm 1:** Robust Levenberg–Marquardt (ECME)

**Require:** Data  $\mathbf{y}$ , mapping  $\mathbf{f}(\boldsymbol{\theta})$ , Jacobian  $\mathbf{J}(\boldsymbol{\theta})$ ,  $\nu$ , initial value  $\boldsymbol{\theta}^0$

```

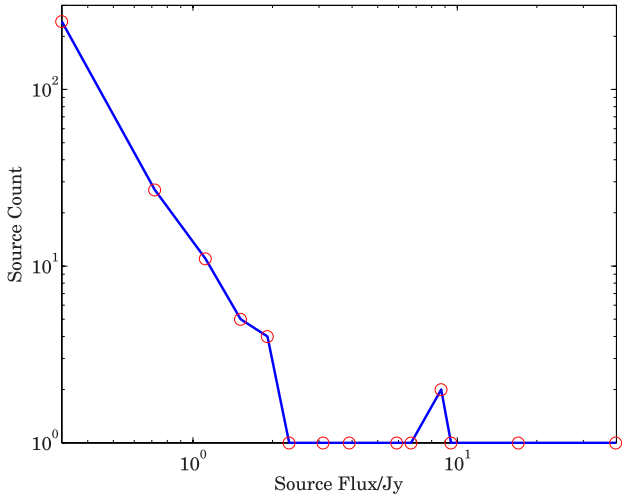
1:  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^0$ ;  $w_i \leftarrow 1$ ;  $\lambda \leftarrow 1$ 
2: while  $l < \max$  EM iterations do
3:    $k \leftarrow 0$ ;  $\eta \leftarrow 2$ 
4:    $\mathbf{J}(\boldsymbol{\theta}) \leftarrow \mathbf{WJ}(\boldsymbol{\theta})$ 
5:    $\mathbf{A} \leftarrow \mathbf{J}(\boldsymbol{\theta})^T \mathbf{J}(\boldsymbol{\theta})$ ;  $\mathbf{e} \leftarrow \mathbf{W}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$ ;  $\mathbf{g} \leftarrow \mathbf{J}(\boldsymbol{\theta})^T \mathbf{e}$ 
6:   found  $\leftarrow (\|\mathbf{g}\|_\infty < \epsilon_1)$ ;  $\mu \leftarrow \tau \max \mathbf{A}_{ii}$ 
7:   while (not found) and ( $k < \max$  iterations)
8:      $k \leftarrow k + 1$ ; Solve  $(\mathbf{A} + \mu \mathbf{I})\mathbf{h} = \mathbf{g}$ 
9:     if  $\|\mathbf{h}\| < \epsilon_2(\|\boldsymbol{\theta}\| + \epsilon_2)$ 
10:       found  $\leftarrow \text{true}$ 
11:   else
12:      $\boldsymbol{\theta}_{new} \leftarrow \boldsymbol{\theta} + \mathbf{h}$ 
13:      $\rho \leftarrow (\|\mathbf{e}\|^2 - \|\mathbf{W}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}_{new}))\|^2) / (\mathbf{h}^T (\mu \mathbf{h} + \mathbf{g}))$ 
14:     if  $\rho > 0$  then
15:        $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_{new}$ 
16:        $\mathbf{J}(\boldsymbol{\theta}) \leftarrow \mathbf{WJ}(\boldsymbol{\theta})$ 
17:        $\mathbf{A} \leftarrow \mathbf{J}(\boldsymbol{\theta})^T \mathbf{J}(\boldsymbol{\theta})$ ;  $\mathbf{e} \leftarrow \mathbf{W}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$ ;  $\mathbf{g} \leftarrow \mathbf{J}(\boldsymbol{\theta})^T \mathbf{e}$ 
18:       found  $\leftarrow (\|\mathbf{g}\|_\infty \leq \epsilon_1)$ 
19:        $\mu \leftarrow \mu \max(1/3, 1 - (2\rho - 1)^3)$ ;  $\eta \leftarrow 2$ 
20:   else
21:      $\mu \leftarrow \mu \eta$ ;  $\eta \leftarrow 2\eta$ 
22:   end if
23: end while
24: end while
25: Update weights  $w_i \leftarrow \lambda \frac{\nu+1}{\nu + (y_i - \mathbf{f}_i(\boldsymbol{\theta}))^2}$ 
26: Update  $\lambda \leftarrow \frac{1}{N} \sum_{i=1}^N w_i$ 
27: Update  $\nu$  using (24)
28:  $l \leftarrow l + 1$ 
29: end while
30: return  $\boldsymbol{\theta}$ 

```

## 5 SIMULATION RESULTS

In this section, we provide results based on simulations to convince the robustness of our proposed calibration approach. We simulate an interferometric array with  $R = 47$  stations, with the longest baseline of about 30 km. We simulate an observation centred at the north celestial pole, with a duration of 6 h at 150 MHz. The integration time for each data sample is kept at 10 s. For the full duration of the observation, there are 2160 data points. Each data point consists of 1081 baselines and 8 real values corresponding to the  $2 \times 2$  complex visibility matrix.

The sky is simulated to have 300 sources, uniformly distributed over a field of view of  $12^\circ \times 12^\circ$ . The intensities of the sources are drawn using a power-law distribution, with the peak intensity at 40 Jy. In Fig. 3, we show the histogram of the intensities of the sources. Our intention is to compare the fluxes of the weak sources, i.e. the sources with intensities less than or equal to 1 Jy, after directional calibration is performed. In order to do that, we corrupt the visibilities of the bright sources with directional errors that vary slowly with time. We consider three scenarios here: we only corrupt the signals of the sources that have intensities greater than (i) 1, (ii) 2 and (iii) 5 Jy. For the simulated sky model, there are 28, 11 and 7 sources that have fluxes greater than 1, 2 and 5 Jy, respectively.



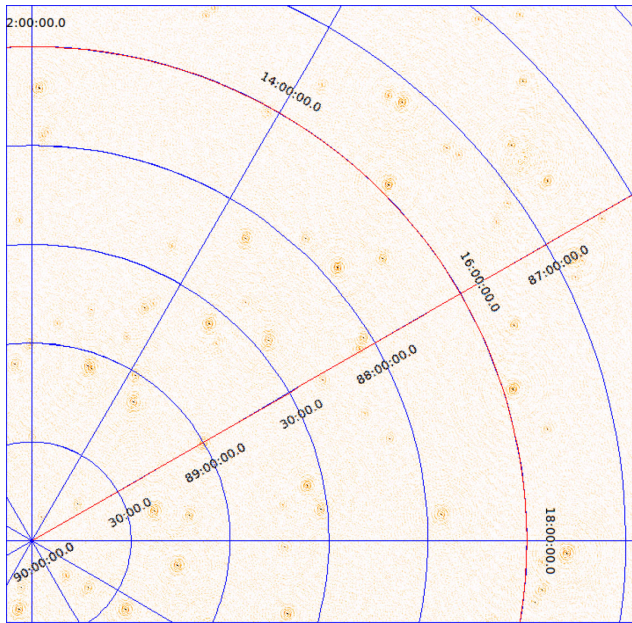
**Figure 3.** Histogram of the fluxes of the 300 simulated sources. The peak flux is 40 Jy.

Note that in each case, we do not corrupt the signals of the weak sources as our only objective is to find the recovered flux after directional calibration and subtraction of the bright sources from the data, although in reality all sources will be corrupted by similar directional errors. Finally, we add zero mean white Gaussian noise to the simulated data, with the signal-to-noise ratio (SNR) defined as

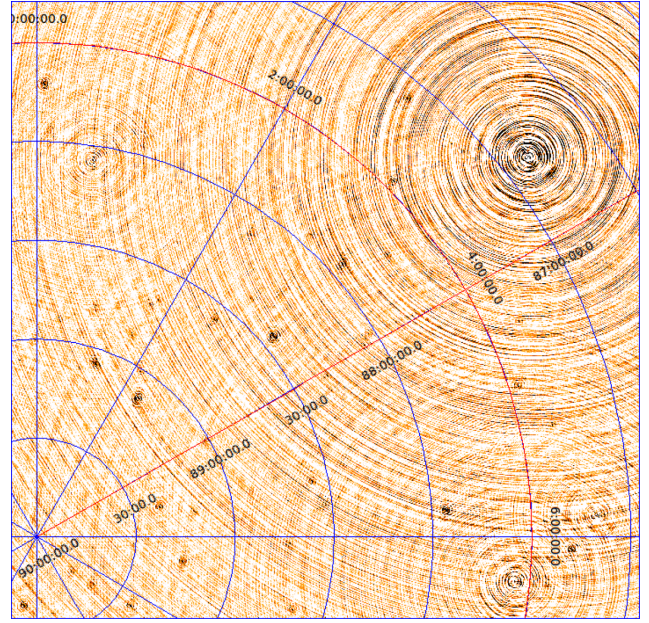
$$\text{SNR} \triangleq 10 \log_{10} \left( \frac{\sum_{p,q} \|\mathbf{V}_{pq}\|^2}{\sum_{p,q} \|\mathbf{N}_{pq}\|^2} \right) \text{ dB}. \quad (26)$$

In all simulations, we have kept the SNR at 5 dB.

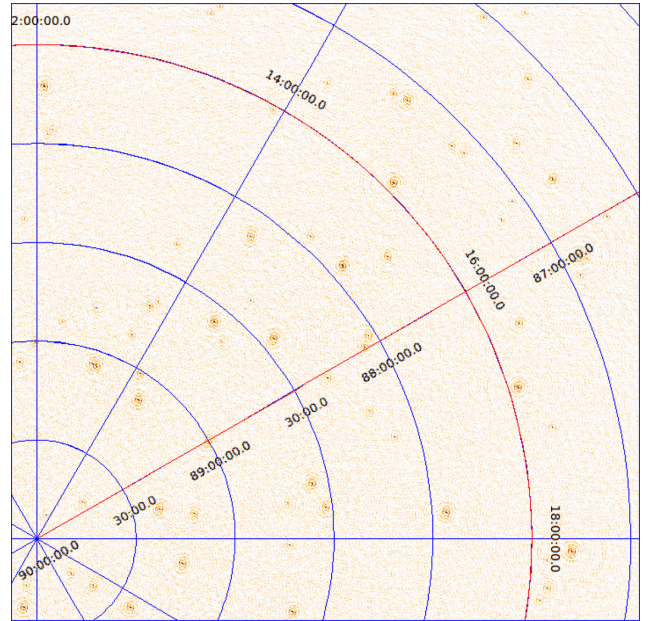
In Fig. 4, we show some of the weak sources (with intensities less than 1 Jy) over a  $4^\circ \times 4^\circ$  area of the field of view. In Fig. 5, we have also added the bright sources with slowly varying directional



**Figure 4.** Simulated image of  $4^\circ \times 4^\circ$  of the sky, showing only weak sources with intensities less than 1 Jy.



**Figure 5.** Simulated image of the sky where bright sources with fluxes greater than 1 Jy have been corrupted with directional errors. Due to these errors, there are artefacts throughout the image that make it difficult to study the fainter background sources.

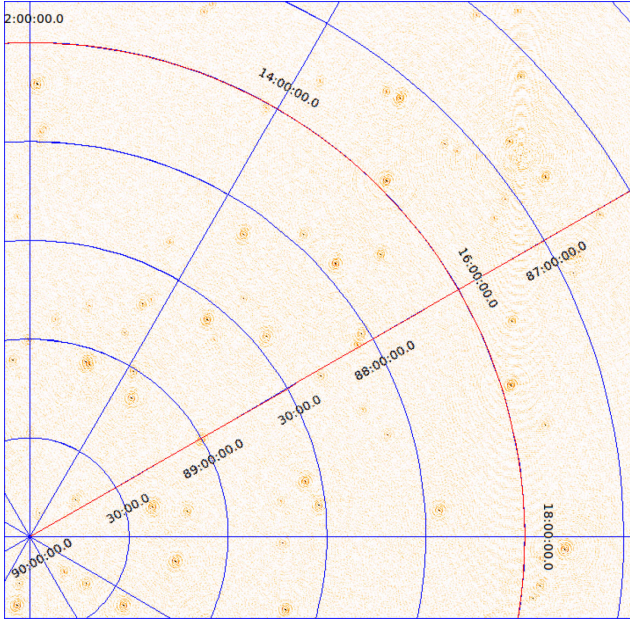


**Figure 6.** Image of the sky where the bright sources have been calibrated and subtracted from the data to reveal the fainter background sources. The traditional calibration based on a Gaussian noise model is applied.

errors. Note that in order to recover Fig. 4 from Fig. 5, directional calibration is essential.

In Fig. 6, we show the image after directional calibration along the bright sources and subtraction of their contribution from the data, using traditional calibration based on a Gaussian noise model. On the other hand, in Fig. 7, we show the image after directional calibration and subtraction using a robust noise model. With respect to the subtraction of the bright sources from the data, both normal





**Figure 7.** Image of the sky where the bright sources have been calibrated and subtracted from the data to reveal the fainter background sources. The robust calibration proposed in this paper is applied.

calibration and robust calibration show equal performance as seen from Figs 6 and 7.

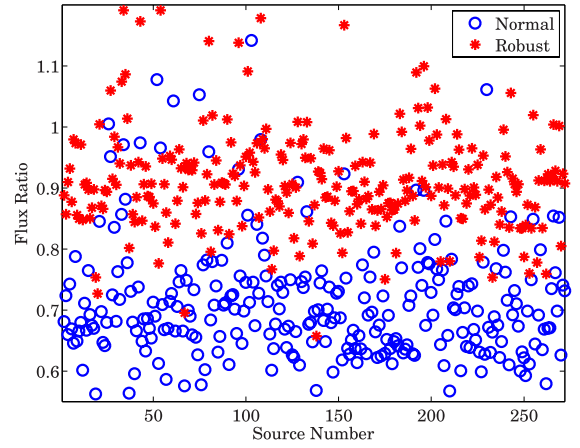
We perform Monte Carlo simulations with different directional gain and additive noise realizations for each scenario (i), (ii) and (iii) as outlined previously. For each realization, we image the data after subtraction of the bright sources and compare the flux recovered for the weak sources before and after directional calibration. The directional calibration is performed for every 10 time samples (every 100 s duration). Therefore, the number of data points used for each calibration ( $N$ ) is  $10 \times 1081 \times 8 = 86480$  and the number of real parameters estimated are  $47 \times 8 \times 28 = 10528$ ,  $47 \times 8 \times 11 = 4136$  and  $47 \times 8 \times 7 = 2632$ , respectively, for scenarios (i), (ii) and (iii). For each scenario (i), (ii) and (iii), we perform 100 Monte Carlo simulations.

Our performance metric is the ratio between the recovered peak flux of the weak sources and the original flux of each source. We calculate the average ratio (recovered flux/original flux) over all Monte Carlo simulations. In Figs 8–10, we show the results obtained for scenarios (i), (ii) and (iii), respectively.

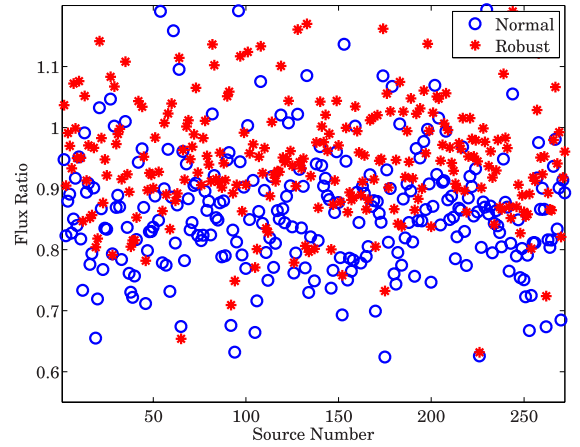
We observe two major characteristics in Figs 8–10. First, we see that as we calibrate over an increasing number of directions (and subtract an increasing number of sources), the recovered flux is reduced. Secondly, in all scenarios, robust calibration recovers more flux compared to normal calibration. To illustrate this point, we also plot in Fig. 11, the ratio between the recovered flux using robust calibration and the recovered flux using normal calibration. As we see from Fig. 11, we almost always get a value greater than 1 for this ratio, indicating that we recover more flux using robust calibration.

We summarize our findings in Table 1. We see that at worst case, the performance of normal calibration gives a flux reduction of about 20 per cent compared to robust calibration.

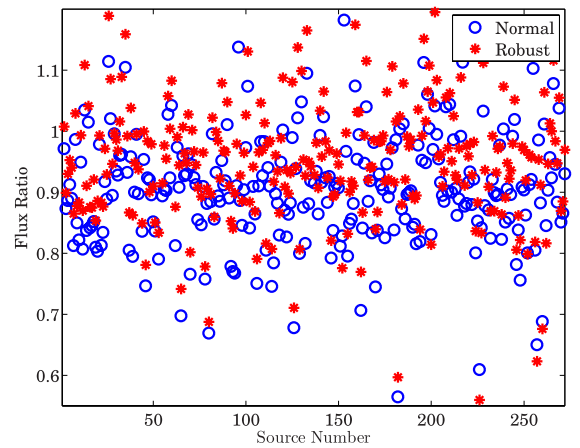
Up to now, we have only considered the sky to consist of only point sources. In reality, there is diffuse structure in the sky. This diffuse structure is seldom incorporated into the sky model during calibration either because it is too faint or because of the



**Figure 8.** Ratio between the recovered flux and the original flux of each of the weak sources, when bright sources ( $>1$  Jy) are subtracted.



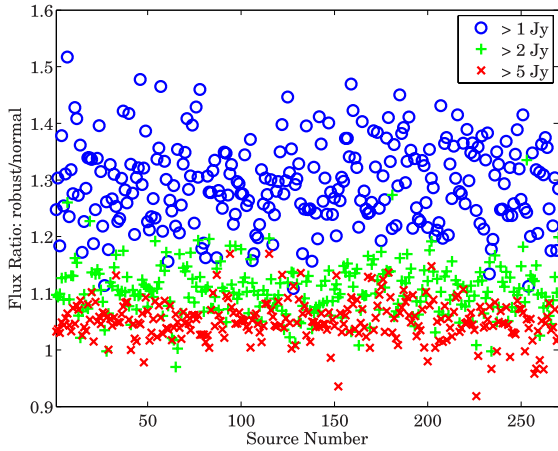
**Figure 9.** Ratio between the recovered flux and the original flux of each of the weak sources, when bright sources ( $>2$  Jy) are subtracted.



**Figure 10.** Ratio between the recovered flux and the original flux of each of the weak sources, when bright sources ( $>5$  Jy) are subtracted.

complexity of modelling it accurately. We have also done simulations where there is faint diffuse structure in the sky and only the bright foreground sources are calibrated and subtracted. We have chosen scenario (i) in the previous simulation except that we have

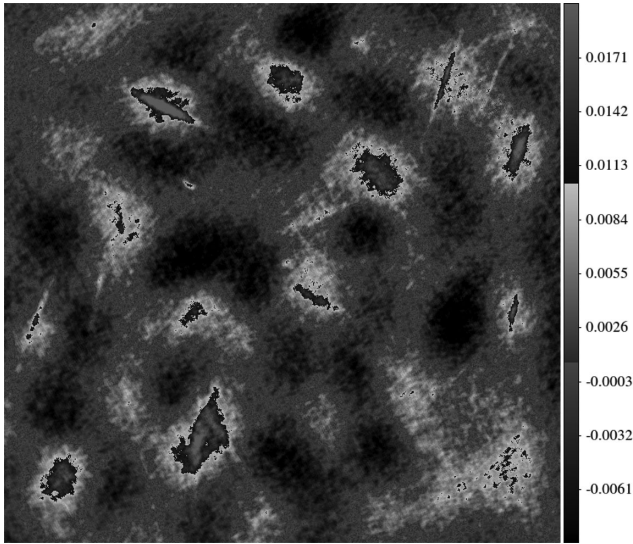




**Figure 11.** Ratio between the recovered flux using robust calibration and the recovered flux using normal calibration. Almost always, robust calibration recovers more flux compared with normal calibration. The different colours indicate different scenarios where the number of bright sources subtracted is varied.

**Table 1.** Comparison of the reduction of flux of the weak background sources with normal and robust calibration.

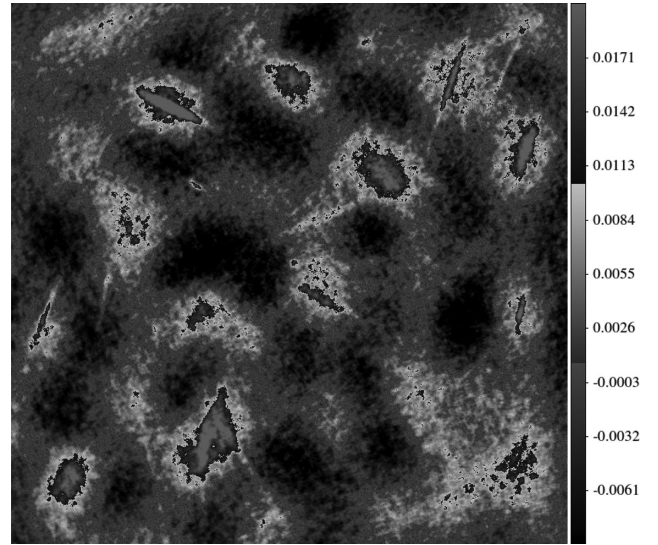
No. of sources calibrated and subtracted	Lowest flux of the subtracted sources (Jy)	Average reduction of the flux of weak background sources (per cent)	
		Normal calibration	Robust calibration
28	1	28.7	8.2
11	2	12.3	3.2
7	5	7.9	3.1



**Figure 12.** The average residual image of the diffuse structure after subtracting the bright sources by normal calibration. The colour scale is in Jy per PSF.

replaced the sources below 1 Jy with Gaussian sources with peak intensities below 1 Jy and with random shapes and orientations.

In Fig. 12, we have shown the residual image of a  $6^\circ \times 6^\circ$  area in the sky after removing all sources brighter than 1 Jy. The residual



**Figure 13.** The average residual image of the diffuse structure after subtracting the bright sources by robust calibration. The colour scale is in Jy per PSF and is the same as in Fig. 12.

image is obtained by averaging 100 Monte Carlo simulations. The equivalent image for robust calibration is given in Fig. 13.

As seen from Figs 12 and 13, there is more flux in the diffuse structure after robust calibration. This is clearly seen in the bottom right-hand corner of both figures, where Fig. 13 has more positive flux than in Fig. 12.

## 6 CONCLUSIONS

We have presented the use of Student's  $t$ -distribution in radio interferometric calibration. Compared with traditional calibration that has an underlying Gaussian noise model, robust calibration using Student's  $t$ -distribution can handle situations where there are model errors or outliers in the data. Moreover, by automatically selecting the number of degrees of freedom  $\nu$  during calibration, we have the flexibility of choosing the appropriate distribution even when no outliers are present and the noise is perfectly Gaussian. For the specific case considered in this paper, i.e. the loss of coherency or flux of unmodelled sources, we have given simulation results that show the significantly improved flux preservation with robust calibration. Future work would focus on adopting this for pipeline processing of massive data sets from new and upcoming radio telescopes.

## ACKNOWLEDGEMENTS

We thank the reviewer, Fred Schwab, for his careful review and insightful comments. We also thank the Editor and Assistant Editor for their comments. We also thank Wim Brouw for commenting on an earlier version of this paper.

## REFERENCES

- Aravkin A., Friedlander M., van Leeuwen T., 2012, Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Robust Inversion Via Semistochastic Dimensionality Reduction. IEEE, Piscataway, NJ, p. 5425
- Bartkowiak A., 2007, Proc. IEEE 6th Int. Conf. on Computer Information Systems and Industrial Management Applications (CISIM'07), Should

- Normal Distribution be Normal? The Student's T Alternative. IEEE, Piscataway, NJ, p. 3
- Boonstra A., van der Veen A., 2003, *IEEE Trans. Signal Process.*, 51, 25
- Cornwell T., Fomalont E. B., 1999 in Taylor G. B., Carilli C. L., Perley R. A., eds, *ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II*. Astron. Soc. Pac., San Francisco, p. 187
- Cornwell T. J., Wilkinson P. N., 1981, *MNRAS*, 196, 1067
- Dempster A., Laird N., Rubin D., 1977, *J. R. Stat. Soc. B*, 39, 1
- Fessler J., Hero A., 1994, *IEEE Trans. Signal Process.*, 42, 2664
- Gosset W. S., 1908, *Biometrika*, 6, 1
- Hamaker J. P., Bregman J. D., Sault R. J., 1996, *A&AS*, 117, 96
- Kazemi S., Yatawatta S., Zaroubi S., Labropoulos P., de Bruyn A., Koopmans L., Noordam J., 2011, *MNRAS*, 414, 1656
- Lange K., Little R., Tylor J., 1989, *J. Am. Stat. Assoc.*, 84, 881
- Levenberg K., 1944, *Q. J. App. Math.*, 2, 164
- Li S., Wang H., Chai T., 2006, *American Control Conference*. Am. Autom. Control Counc., Troy, NY, p. 6
- Liu C., Rubin D., 1995, *Stat. Sin.*, 5, 19
- Lourakis M., 2004, Technical report, Institute of Computer Science, Foundation for Research and Technology, Hellas (FORTH) on Levmar: Levenberg–Marquardt nonlinear least squares algorithms in C/C++.
- Madsen K., Nielsen H., Tingleff O., 2004, *Lecture Notes: Methods for non-linear least squares problems*, 2nd edn. Technical University of Denmark
- Marquardt D., 1963, *SIAM J. App. Math.*, 11, 431
- Martí-Vidal I., Marcaide J., 2008, *A&A*, 480, 289
- Martí-Vidal I., Ros E., Perez-Torres M., Guirado J., Jiménez-Monferrer S., Marcaide J., 2010, *A&A*, 515, A53
- Offringa A., de Bruyn A., Biehl M., Zaroubi S., Bernardi G., Pandey V., 2010, *MNRAS*, 405, 155
- Schwab F., 1982, *VLA Scientific Memorandum*, 136, 1
- Thompson A., Moran J., Swenson G., 2001, *Interferometry and Synthesis in Radio Astronomy*, 3rd edn. Wiley Interscience, New York
- van der Tol S., Jeffs B., van der Veen A., 2007, *IEEE Trans. Signal Process.*, 55, 4497
- van der Veen A., Leshem A., Boonstra A., 2004, *Exp. Astron.*, 17, 231
- Yatawatta S., Kazemi S., Zaroubi S., 2012, *Proc. IEEE Int. Symp. on Signal Processing and Information Technology*

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.